DOCUMENT RESUME

ED 135 860 TM 006 091

AUTHOR Timm, Neil H.
TITLE Growth Curve Analysis.
PUB DATE [Apr 77]
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Analysis of Covariance; *Analysis of Variance; *Hypothesis Testing; *Mathematical Models; Matrices; Research Problems; *Statistical Analysis
IDENTIFIERS *Growth Curve Analysis

ABSTRACT

Several procedures proposed in the literature for the analysis of growth curves are reviewed. Particular attention is given to the current issues in this area to guide practitioners in the selection of the most appropriate methodology. (Author)

GROWTH CURVE ANALYSIS[1]

Neil H. Timm

University of Pittsburgh

Abstract

Several procedures proposed in the literature for the analysis of growth curves are reviewed. Particular attention is given to the current issues in this area to guide practitioners in the selection of the most appropriate methodology.

1. Introduction

In many experimental situations, especially longitudinal growth curve studies, data are collected on several variates and a subject is observed on each of the variates over time. Designs with multivariate observations on p_o variates observed over q time points are often referred to as multi-response repeated measures or growth curve designs.

To analyze multi-response growth curve data obtained on N subjects, the data are conveniently arranged in a data matrix $Y_o(N \times p_o q)$ where the first q columns correspond to variable one, the next q to variable 2, the next q to variable 3 and so on up to the p_o^{th} variable.

Given a data matrix Y_o with N_i subjects in m groups, $\sum_{i=i}^{m} N_i = N$, where for simplicity suppose that three measures are recorded on each subject at q time points so that associated with each subject are 3q measurements all correlated with unknown variance-covariance matrix Σ_o, the i^{th} growth curve for each of the three multivariate responses may be represented by

$$(1.1)\qquad \begin{array}{ll} \beta_{i0} + \beta_{i1}t + \ldots + \beta_{i,p_1-1}t^{p_1-1} & p_1 \le q \\ \theta_{i0} + \theta_{i1}t + \ldots + \theta_{i,p_2-1}t^{p_2-1} & p_2 \le q \\ \xi_{i0} + \xi_{i1}t + \ldots + \xi_{i,p_3-1}t^{p_3-1} & p_3 \le q \end{array}$$

---

[1]Paper presented at the annual meeting of the American Educational Research Association, New York City, April 1977.

Employing matrix notation with the matrices B and P defined:

$$
\underset{m \times (p_1+p_2+p_3)}{B} = \begin{pmatrix}
\beta_{10} & \beta_{11} & \cdots & \beta_{1,p_1-1} & \theta_{10} & \theta_{11} & \cdots & \theta_{1,p_2-1} & \xi_{10} & \xi_{11} & \cdots & \xi_{1,p_3-1} \\
\beta_{20} & \beta_{21} & \cdots & \beta_{2,p_2}{}^{-1} & \theta_{20} & \theta_{21} & \cdots & \theta_{1,p_2-1} & \xi_{20} & \xi_{21} & \cdots & \xi_{2,p_3-1} \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\beta_{m0} & \beta_{m1} & \cdots & \beta_{m,p_1-1} & \theta_{m0} & \theta_{m1} & \cdots & \theta_{m,p_2-1} & \xi_{m0} & \xi_{m1} & \cdots & \xi_{m,p_3-1}
\end{pmatrix}
$$

$$
\underset{(p_1+p_2+p_3) \times 3q}{P} = \begin{pmatrix}
1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
t_1 & t_2 & \cdots & t_q & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
t_1^{p_1-1} & t_2^{p_1-1} & \cdots & t_q^{p_1-1} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
0 & 0 & \cdots & 0 & t_1^{p_2-1} & t_2^{p_2-1} & \cdots & t_q^{p_2-1} & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & t_1 & t_2 & \cdots & 1 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & t_1^{p_3-1} & t_2^{p_3-1} & \cdots & t_q^{p_3-1}
\end{pmatrix}
$$

(1.1) may be represented by the matrix product BP. The matrix P given above has been represented as a super diagonal Vandermode matrix; alternatively, we could have used unnormalized or normalized orthogonal polynomials. Letting X (N x m) denote a design matrix, the growth curve model is represented as

$$(1.2) \qquad \begin{aligned} E\,(Y_o) &= XBP \\ V\,(Y_o) &= I_N \otimes \Sigma_o . \end{aligned}$$

Representing the growth curve model (GCM) as (1.2), some common multi-response hypotheses of the form

$$(1.3) \qquad H_o\colon \; CBA = \Gamma$$

where C (g x m) of full rank g, A ($p_o q$ x u) of full rank u, and Γ(g x u) are known matrices may be tested.

To analyze growth curve data on one variate ($p_o = 1$) observed at q time points using (1.2), most authors reduce the GCM to a standard MANOVA or MANCOVA model. The procedures developed using these approaches and the problems encountered is the major topic of this paper. From the representation of B and P, whether $p_o = 1$ or $p_o > 1$ the extension to more than one variate is immediate.

2. Standard MANOVA Model

In a multivariate experiment involving N subjects on which p measurements are observed, the data obtained may be represented by a (N x p) data matrix Y. Assuming the standard MANOVA model (SMM) describes the experiment, the model for the random matrix Y is denoted as follows.

$$\begin{aligned} E\,(Y) &= XB \\ V\,(Y) &= I_N \otimes \Sigma \end{aligned}$$

The matrix X (N x m) is a known design matrix of rank $r \leq m \leq N$, B (m x p) is a matrix of unknown nonrandom parameters and Σ(p x p) is a positive definite (p.d.) variance-covariance matrix of any p-variate row vector of Y.

In the SMM, we are usually interested in estimating linear parametric functions of the form $\psi = \underset{\sim}{c}'B\underset{\sim}{a}$ and sets of functions of the form CBA where $\underset{\sim}{c}$(m x 1),

$\underset{\sim}{a}$ (p x 1), C (g x m) and A (p x u) are known, the R(C) = g and the R (A) = u. Assuming ψ is estimable, $\underset{\sim}{c}$ belongs to the space spanned by X'X or equivalently $\underset{\sim}{c}' = \underset{\sim}{c}'H$ where $H = (X'X)^{-}(X'X)$, Rao (1973) and Roy (1975) have shown that the best linear unbiased estimator (BLUE) of $\psi = \underset{\sim}{c}' B \underset{\sim}{a}$ when it is estimable, is given by

$$\hat{\psi} = \underset{\sim}{c}' \hat{B} \underset{\sim}{a} \qquad (2.2)$$
$$\hat{B} = (X'X)^{-} X'Y$$

The variance of $\hat{\psi}$ is

$$(\underset{\sim}{a}' \Sigma \underset{\sim}{a}) (\underset{\sim}{c}'(X'X)^{-} \underset{\sim}{c})$$

(see, e.g. Timm, 1975, Section 3.6). Considering the linear set CBA, Roy (1964) showed that if CB_oA is any other unbiased estimator of CBA, other than $C\hat{B}A$, that

$$V(CB_oA) - V(C\hat{B}A) \text{ is (p.s.d)}$$
$$ch_{max}[V(CB_oA)] \geq ch_{max}[V(C\hat{B}A)]$$
$$Tr\,[V(CB_oA)] > [Tr\ V(C\hat{B}A)]$$
$$|V(CB_oA)| > |V(C\hat{B}A)|$$

where the $V(C\hat{B}A) = C(X'X)^{-}C' \otimes A' \Sigma A$.

Hence, $\hat{\Psi} = C\hat{B}A$ is a unique solution in terms of the minimization of the trace criterion, and the generalized variance criterion because of the strict inequality. Although $\hat{\Psi}$ yields a minimum in terms of the maximum root criterion, the solution is not unique.

The estimation of Ψ under 2.1, the multivariate Gauss-Markoff setup, does not require distributional assumptions about the row vectors of Y. If, however, we assume that each of the rows of Y are independently normally distributed, then maximum likelihood estimators of the functions of B may be obtained. Writing the likelihood function as

$$(2.4) \quad (2\pi)^{-Np/2} \, |\Sigma|^{-N/2} \exp\left[-\frac{1}{2} \operatorname{tr} \Sigma^{-1} (Y-XB)'(Y-XB) \right]$$

and solving the likelihood equations, the maximum likelihood estimators (MLE's) of B and Σ are:

$$(2.5) \quad \begin{aligned} \hat{B} &= (X'X)^{-}X'Y \\ \hat{\Sigma} &= Y'(I - X(X'X)^{-}X')\,Y/N \end{aligned}$$

Hence, the MLE of $\widehat{CBA}$ is identical to the BLUE. $\hat{\Sigma}$ is the unique MLE of Σ ; to obtain an unbiased estimator, $\hat{\Sigma}$ is multiplied by $N/(N-r)$.

Hypothesis testing under 2.1, assuming normality, takes the general form

$$(2.6) \qquad H_o : \; CBA = \Gamma$$

where C $(g \times m)$ is a known matrix of rank $g \le m$, A $(p \times u)$ is a known matrix of rank $u \le p$, $\Gamma\,(g \times u)$ is a known matrix (usually a zero matrix) and CBA is estimable. Defining hypothesis and error sums of squares and products matrices of the form

$$(2.7) \quad \begin{aligned} Q_h &= (C\hat{B}A - \Gamma)'\,(C(X'X)^{-}C')^{-1}\,(CBA - \Gamma) \\ Q_e &= A'Y\,(I - X(X'X)^{-}X')YA \end{aligned}$$

and

$$\lambda_i = ch_i(Q_h Q_e^{-1}), \; i = 1, 2, .., s$$

where $s = \min(R(C), R(A))$, numerous test criteria have been proposed to test H_o. They are:

Wilks' Lambda Criterion -

$$\frac{|Q_e|}{|Q_n + Q_e|} = \prod_{i=1}^{s} (1 + \lambda_i)^{-1}$$

Hotelling's Trace Criterion -

$$\mathrm{Tr}\ (Q_h Q_e^{-1}) = \sum_{i=1}^{s} \lambda_i$$

Roy's Largest Root Criterion -

$$ch_1\ (Q_h Q_e^{-1}) = \lambda_1 \text{ or } ch_1\ (Q_h\ (Q_h + Q_e)^{-1}) = \theta_1 = \frac{\lambda_1}{1+\lambda_1}$$

Pillai's Trace Criterion -

$$\mathrm{Tr}\ (Q_h(Q_h+Q_e)^{-1}) = \sum_{i=1}^{s} \theta_i = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$$

Another criterion, although not extensively tabled includes:

Roy's Minimum Root Criterion - $ch_s(Q_h Q_e^{-1}) = \lambda_s$

Although no one criterion is uniformly most powerful, the studies of Schatzoff (1966) and Olson (1974) show that under normality Roy's criterion is best for certain restrictive alternatives and that Wilks' criterion is best for a wide class of alternatives. The best robust criterion appears to be Pillai's trace criterion.

3. Potthoff and Roy Model

While the SMM is applicable in many experimental situations, the model has several limitations if an experimenter wants to analyze and fit growth curves to data collected over time. To analyze data obtained from a growth curve experiment, Potthoff and Roy (1964) developed the growth curve model (GCM) which is a simple extension of the SMM.

The model considered by Potthoff and Roy is given by

$$\begin{aligned} E(Y_o) &= X\ B\ P \\ V(Y_o) &= I_N \otimes \Sigma_o. \end{aligned} \tag{3.1}$$

where Y_o (N x q) is a data matrix, X (N x m) is a known design matrix, B (m x p) is a matrix of unknown nonrandom parameters, P (p x q) is a known matrix of full rank $p \leq q$, Σ_o (q x q) is p.d. and the rows of Y_o are independently normally distributed

Comparing the GCM with the SMM, we see that only the post matrix P has been added to the model. This implies that each response variate can be expressed as a linear regression model of the form

$$E(\underset{\sim}{y}_i) = P' \underset{\sim}{\beta}_i$$

where $\underset{\sim}{y}_i$ $(q \times 1)$ is the observation vector for the i^{th} subject and $\underset{\sim}{\beta}_i$ is a vector of unknown parameters.

To analyze (3.1), Potthoff and Roy suggested the transformation

$$(3.2) \qquad Y = Y_o G^{-1} P' (P G^{-1} P')^{-1}$$

where $G(q \times q)$ is any symmetric positive definite weight matrix either non-stochastic or independent of Y_o such that $PG^{-1}P'$ is of full rank. Employing the transformation in (3.2), the matrix $Y(N \times p)$ will be distributed mutually independently normal with unknown p.d. variance-covariance matrix

$$\underset{(p \times p)}{\Sigma} = [P(G')^{-1}P']^{-1} P(G')^{-1} \Sigma_o G^{-1} P' (P G^{-1} P')^{-1}$$

and mean $E(Y) = XB$. Hence, by using (3.2) we have reduced the GCM to the SMM with minor limitations on the selection of G.

Motivation for the selection of the transformation in (3.2) by Potthoff and Roy is contained in Appendix B of their (1964) paper; they show that the BLUE of an estimable linear parametric function $\psi = \underset{\sim}{c}'B\,\underset{\sim}{a}$ (where the estimability conditions are that $\underset{\sim}{c}$ belongs to the space spanned by X'X and $\underset{\sim}{a}$ belongs to the space spanned by the columns of P) is given by

$$(3.3) \qquad \begin{aligned} \hat{\psi} &= \underset{\sim}{c}'\hat{B}\,\underset{\sim}{a} \\ \hat{B} &= (X'X)^{-} X' Y_o \Sigma_o^{-1} P' (P \Sigma_o^{-1} P')^{-1} \end{aligned}$$

Since (3.1) reduces to (2.1) under (3.2), we see that upon substituting Y in (3.2) into (2.2) that

$$\hat{B} = (X'X)^{-}X'Y_oG^{-1}P'(PG^{-1}P')^{-1},$$

with G replacing Σ_o in (3.3), is very close to the BLUE.

To test hypotheses of the form

(3.4) $\quad H_o: \quad CBA = \Gamma$

under (3.1), we merely have to substitute Y defined in (3.2) into the expression for Q_h and Q_e in (2.6). The degrees of freedom for the hypotheses is $\nu_h = R(C)=g$ and the degrees of freedom for error is $\nu_e=N-R(X)=N-r$.

Setting $\Gamma=0$ in (3.4) and letting Y be defined as in (3.2), the hypotheses and error sum of square and products matrices take the following form.

$$(3.5) \quad \begin{aligned} Q_h &= A'Y'X(X'X)^{-}C'(C(X'X)^{-}C')^{-1}C(X'X)^{-}X'YA \\ Q_e &= A'Y'(I-X(X'X)^{-}X')YA \end{aligned}$$

where $\nu_h = g$ and $\nu_e = N-r$.

Under the SMM, we said that no criteria is uniformly most powerful. This is also the case for the GCM; however, in the GCM we have the additional problem of selecting the weight matrix G when $p < q$. If $p=q$, the transformation in (3.2) reduces to

$$Y = Y_oP^{-1}$$

or if P is an orthogonal matrix so that $P^{-1} = P'$,

$$Y=Y_oP'$$

and there is no need to choose G. This was the approach taken by Bock (1963a) and the one used in the development of the NYMBUL package, Bock (1963b) and Finn (1972). If $p < q$ the choice of G is important since it affects the variance of $\hat{\psi}$ which increases as G^{-1} departs from Σ_o^{-1}, the power of the tests and the widths of confidence bands.

A simple choice of G is to set G=I. Then

$$Y=Y_oP'(PP')^{-1}$$

Such a choice of G will certainly simplify one's calculations; however, it is not the best choice in terms of power since information is lost by reducing

Y_o to Y unless G is set equal to Σ_o. The estimator of $\hat{\psi} = \underset{\sim}{c}'\hat{B}\underset{\sim}{a}$, when it is estimable and G is set equal to I, is the BLUE of ψ assuming $\Sigma_o = \sigma^2 I$.

4. Rao-Khatri Conditional Model

To try to avoid the arbitrary choice of the matrix G in Potthoff and Roy's model and its effect on estimates and tests, Rao (1965, 1966, 1967, 1972) and Khatri (1966) independently developed an alternative reduction of model (3.1) to a conditional model.

$$(4.1) \quad E(Y|Z) = XB+Z\Gamma$$

where Y (N x p) is a data matrix, X(N x m) is a known design matrix, B(m x p) is a matrix of unknown nonrandom parameters, Z(N x h) is a matrix of covariates and Γ (h x p) is a matrix of unknown regression coefficients.

To reduce (3.1) to (4.1) a q x q nonsingular matrix $H = (H_1 H_2)$ is constructed so that the columns of H_1 form a basis for the vector space spanned by the rows of P, $PH_1 = I$ and $PH_2 = 0$. When the rank of P is p, H_1 and H_2 can be selected as

$$H_1 = G^{-1}P'(PG^{-1}P')^{-1} \qquad H_2 = I-H_1P$$

where G is an arbitrary positive definite matrix. Such a matrix H is not unique; however, estimates and tests are invariant for all choices of H satisfying the specified conditions (see Khatri, 1966). Hence, G in the expression for H_1 does not affect estimates or tests under (4.1). By setting

$$(4.2) \quad \begin{aligned} Y &= Y_oH_1 = Y_oG^{-1}P'(PG^{-1}P')^{-1} \\ Z &= Y_oH_2 \end{aligned}$$

$E(Y) = XB$ and $E(Z) = 0$; thus, the expected value of Y given Z is seen to be of the form specified in (4.1), Khatri (1966) and Grizzle and Allen (1969). Using (4.1), the information contained in the covariates $Z = Y_oH_2$, which is ignored in the Potthoff-Roy reduction, is utilized.

Both Rao and Khatri argued that the BLUE under the conditional model of $\psi = \underset{\sim}{c}'B\underset{\sim}{a}$ is more efficient than that obtained by Potthoff and Roy since their estimator includes information in Z ignored by Potthoff and Roy. This is not the case. As shown by Lee (1974) and Timm (1975) employing the standard multivariate analysis of covariance (MANCOVA) model,

$$\hat{B} = (X'X)^{-}X'Y_oS^{-1}P'(PS^{-1}P')^{-1} \tag{4.3}$$

where $S = Y_o'(I-X(X'X)^{-}X')Y_o$. Khatri (1966) using the maximum likelihood procedure obtained the same result for $\hat{B}$. Thus, if $p< q$, Rao's procedure using q-p covariates, Khatri using maximum likelihood methods and Potthoff and Roy's method weighting by $G^{-1}=S^{-1}$ are identical. Setting G=I in the Potthoff and Roy method is equivalent to not including any covariates in the Rao-Khatri reduction. When p=q, H_2 does not exist; thus, the Rao-Khatri model is not applicable.

Testing the hypothesis

$$H_o: \quad CBA = \Gamma$$

where $\Gamma = 0$, is not the same under the Potthoff and Roy and Rao-Khatri reductions. Employing the standard MANCOVA model,

$$Q_h = A'Y'X(X'X)^{-}C'(CRC')^{-1}C(X'X)^{-}X'YA$$

$$Q_e = A'(PS^{-1}P')^{-1}A \tag{4.4}$$

where

$$R = (X'X)^{-}+(X'X)^{-}X'Y_o(S^{-1}-S^{-1}P'(PS^{-1}P')^{-1}PS^{-1})Y_oX(X'X)^{-}$$

$$Y = Y_oS^{-1}P'(PS^{-1}P')^{-1}$$

$\nu_h = g$, $\nu_e = N-r-h$ and $h=q-p$.

Although Potthoff and Roy's approach does not allow G to be stochastic unless it is independent of Y_o, it is interesting to compare (3.5) and (4.4) if G=S. Then

$$
\begin{aligned}
Q_e &= A'Y(I-X(X'X)^{-}X')YA \\
&= A'(PS^{-1}P')^{-1}PS^{-1}Y_o(I-X(X'X)^{-}X')Y_oS^{-1}P'(PS^{-1}P')^{-1}A \\
&= A'(PS^{-1}P')^{-1}P'(PS^{-1}P')^{-1}A \\
&= A'(PS^{-1}P')^{-1}A
\end{aligned}
$$

which except for the degrees of freedom for error is identical to Q_e obtained under the Rao-Khatri reduction. The sum of squares and products matrix Q_h, however, is not the same.

The development of the GCM by Potthoff and Roy and the subsequent Rao-Khatri reduction has caused a great deal of confusion among experimentors trying to use the model in growth curve studies. The first major paper which helped to clarify and unify the methodologies was by Grizzle and Allen (1969). They also develop a procedure for selecting only a subset of the q-p covariates.

5. Standard GCM

Potthoff and Roy's analysis of the GCM was developed by introducing the transformation

$$Y = Y_oG^{-1}P'(PG^{-1}P')^{-1}$$

to reduce the GCM to the SMM. To avoid having a test procedure that was dependent on an arbitrary positive definite matrix G, Rao (1965) and Khatri (1966) proposed an alternative reduction to the standard MANCOVA model which did not depend on G. Their procedure, as discussed by Grizzel and Allen (1969), depends on selecting the "best" set of q-p covariates. In addition, one may question the use of co-variates that are part of the transformed variables of the dependent variables being analyzed. To avoid these problems, Tubbs, Lewis and Duran (1975) developed a test procedure to test

$$H_o: CBA = \Gamma$$

employing maximum likelihood methods directly under the GCM.

Under the GCM, the maximum likelihood estimator of B is

$$\hat{B} = (X'X)^{-1}X'Y_oS^{-1}P'(PS^{-1}P')^{-1} \tag{5.1}$$

and under H_o $CBA = \Gamma$,

$$(5.2)\quad \hat{B}_{H_o} = \hat{B} - (X'X)^{-}C'(C(X'X)^{-}C')^{-1}(C\hat{B}A - \Gamma)(A'(PS^{-1}P)^{-1}A)^{-1}A'(PS^{-1}P')^{-1}$$

Using the likelihood ratio criterion due to W

$$Q_h = (C\hat{B}A-\Gamma)'\ (C(X'X)^{-}C')^{-} \qquad)$$

$$(5.3)\qquad Q_e = A'(PS^{-1}P')^{-1}A$$

where $\nu_h = g$ and $\nu_e = N-r$.

Comparing this result with that proposed by Rao and Khat see that each Q_h is different, but have the same degrees of freedom and - Q_e is identical for both procedures, but have different degrees of freedom. However, as pointed out by Kleinbaum (1973), both procedures are asymptotically equivalent since they have the same asymptotic Wishart distributions. No information is available about the two procedures for small samples or about the relative power of each procedure.

6. Kleinbaum's Generalized GCM

In the analysis of growth curve data, observations at some time points are missing either by chance or design so that each dependent variate is not measured on each subject. In addition, the design matrix X may not be the same for each dependent variate. While these problems have been discussed in the literature by Trawinski and Bargmann (1964), Srivastava and Roy (1965) and Srivastava (1966, 1967, 1968), extending the theory of the SMM, Kleinbaum (1973) developed a generalized growth curve model (GGCM) for estimating and testing hypothesis when observations are missing either by chance or design with different design matrices corresponding to different response variates.

To develop the GGCM, we assume we have N subjects with observations taken at q time points. Because of the incomplete data, the N subjects are divided into s disjoint sets $S_1, S_2, \ldots, S_s$ where S_i contains N_i subjects. For $i=1,2,\ldots,s$

measurements are obtained at $q_i < q$ time points. Letting Θ_i $(q \times q_i)$ represent an indicator matrix which specifies at which time points data in the set S_i are obtained, the GGCM is represented as

$$(6.1) \qquad \begin{aligned} E(Y_{oi}) &= X_i B P \Theta_i \\ V(Y_{oi}) &= I_{N_i} \otimes \Theta_i' \Sigma_o \Theta_i \end{aligned} \qquad i=1,2,\ldots,s$$

where $Y_{oi}(N_i \times q_i)$ is the data matrix for the i^{th} set S_i, $X_i(N_i \times m)$ is the design matrix for set S_i, $B(m \times p)$ is the matrix of unknown parameters, P is a known matrix of full rank p and Θ_i $(q \times q_i)$ is an indicator matrix of 0's and 1's for the set S_i. Analogous to the GCM, we assume that Y_{oi} and $Y_{oi}'(i \neq i')$ are independent and the rows of Y_{oi} are normally distributed.

Implicit in 6.1 is the fact that each subject in the experiment is from the same family so that $P_i = P\,\Theta_i$ and

$$E(\underset{\sim}{Y}_{ij}) = P_i' \underset{\sim}{\beta}_{ij}$$

where $\underset{\sim}{Y}_{ij}$ $(q_i \times 1)$ is the vector valued observation of the j^{th} subject from set S_i.

As discussed by Srivastava (1967) and more generally by Kleinbaum (1970), to obtain BLUE of every parametric function $\psi = \underset{\sim}{c}'\,B\,\underset{\sim}{a}$ in complex multivariate linear models (linear models with design matrices that are not the same for each dependent variate) that are independent of the unknown elements of the variance-covariance matrix, requires additional restrictive conditions on the model (see, e.g. Kleinbaum, 1970, p. 58). This led Kleinbaum (1973) to consider Best Asymptotically Normal (BAN) estimators for the GGCM which use consistent estimators of Σ_o and generally yield nonlinear estimators with variances that are in large samples the minimum that could be achieved by linear estimators if Σ_o were known.

To obtain a BAN estimator for an estimable function $\psi = \underset{\sim}{c}'B\underset{\sim}{a}$ in the GGCM, (6.1) is conveniently represented in vector notation as

$$E(\underset{\sim}{y}^*_{oi}) = (\Theta_i' P \otimes X_i)\underset{\sim}{\beta}^*$$

$$i=1,2,\ldots,s$$

$$V(\underset{\sim}{y}^*_{oi}) = \Theta_i' \Sigma_o \Theta_i \otimes I_N$$

where $\underset{\sim}{y}^*_{oi}$ $(N_i q_i x1)$ is obtained from Y_{oi} by rolling out Y_{oi} columnwise and $\underset{\sim}{\beta}^*$(mp x 1) is the columnwise version of B. Applying the general univariate least squares theory to (6.2), with $\hat{\Sigma}_o$ any consistent estimator of Σ_o, a BAN estimator of ψ is $\hat{\psi} = \underset{\sim}{c}'\hat{B}\underset{\sim}{a}$ where

$$\hat{\underset{\sim}{\beta}}^* = \sum_{i=1}^{s} P\,\Theta_i((\Theta_i' \hat{\Sigma}_o \Theta_i')^{-1} \Theta_i' P' \otimes X_i' X_i)^{-} \sum \Theta_i (\Theta_i' \hat{\Sigma}_o \Theta_i)^{-1} \otimes X_i')\underset{\sim}{y}^*_{oi}$$

and $\hat{\underset{\sim}{\beta}}^*$ is the columnwise representation of $\hat{B}$.

To test the null hypothesis H_o: CBA = Γ using the GGCM, Kleinbaum (1973) proposes the Wald statistics,

$$W_N = (C'\hat{B}A-\Gamma)' \left(C'\sum_{i=1}^{s} P\Theta_i (\Theta_i' (\Theta_i'\hat{\Sigma}_o\Theta_i)^{-1} \Theta_i P_i' \otimes X'X_i)^{-} C\right)^{-1} (C'\hat{B}\,A-\Gamma)$$

and

$$W_N^* = (C'\hat{B}\,A-\Gamma)' \left(C' (\sum_{i=1}^{s} P\Theta_i (\Theta_i'\hat{\Sigma}_o\Theta_i)^{-1}\Theta_i P' \otimes Q_i)^{-} C\right)^{-1} (C\,\hat{B}\,A-\Gamma)$$

where $Q_i = X_i'X_i - X_i'Y_{oi}F_i'(F_iY_{oi}'Y_{oi}F_i')^{-1} F_iY_{oi}'X_i$

and F_i is a column basis for $I-P_i'(P_iP_i')^{-1}P_i$ so that $P_iF_i' = 0$.

Comparing the Wald Statistics with the test procedures proposed employing the GCM, W_N/N is equal to the LRT procedure and W_N^*/N is equal to the Rao-Khatri method using the Lawley-Hotelling Trace Criterion.

7. Summary

To test hypothesis of the form H_o: CBA=0 assuming a GCM with $p < q$, three approaches have been suggested to applied researchers over the past decade.

<u>Potthoff and Roy</u> -

Using the transformation $Y = Y_o G^{-1}P'(PG^{-1}P')^{-1}$ and forming the estimator $\hat{B} = (X'X)^{-}X'Y$, the hypothesis and error matrices are formed:

$$Q_h = A'Y'X'(X'X)^{-}C'(C(X'X)^{-}C')^{-1}C(X'X)^{-}X'YA$$

$$Q_e \quad [illegible] \; (I-X(X \; [illegible] \;)^{-}X')YA$$

where $\nu_h = g = R(C)$, $\nu_e = N-r$ and G is any symmetric p.d. weight matrix either non-stochastic or independent of Y_o such that $PG^{-1}P'$ is of full rank.

<u>Tubbs, Lewis and Duran</u> -

Using maximum likelihood procedures, which is equivalent to setting G=S in the Potthoff and Roy model, they obtain

$$\hat{B} = (X'X)^{-}X'Y_o S^{-1}P'(PS^{-1}P')^{-1} = (X'X)^{-}X'Y$$

$$Q_h = A'YX(X'X)^{-}C'(C(X'X)^{-}C')^{-1}C(X'X)^{-}X'YA$$

$$Q_e = A'Y'(I-X(X'X)^{-}X')YA = A'(PS^{-1}P')^{-1}A$$

where $\nu_h = g = R(C)$, $\nu_e = N-r$, $S = Y_o'(I-X(X'X)^{-}X')Y_o$ and $Y = Y_o S^{-1}(PS^{-1}P')^{-1}$

<u>Rao-Khatri</u> -

Using a conditional model with

$$\hat{B} = (X'X)^{-}X'Y_o S^{-1}P'(PS^{-1}P')^{-1} = (X'X)^{-}X'Y$$

$$Y = Y_o S^{-1}P'(PS^{-1}P')^{-1}$$

$$S = Y_o'(I-X(X'X)^{-}X')Y_o$$

the matrices

$$Q_h = A'Y'X(X'X)^{-}C'(CRC')^{-1}C(X'X)^{-}X'YA$$

$$Q_e = A'(PS^{-1}P')^{-1}A$$

$$R \quad (X'X)^{-}+(X'X)^{-}X'Y_o(S^{-1}-S^{-1}P'(PS^{-1}P')^{-1}PS^{-1})Y_o X(X'X)^{-}$$

are formed where $\nu_h = g = R(C)$ and $\nu_e = N-r-q+p$.

While the procedure set forth by Khatri has been "accepted" as the usually procedure employed in growth curve studies over the years and is asymptotically equivalent to the procedure proposed by Tubbs, Lewis and Duran, we do not know which of the procedures are best in small samples. Perhaps the determination cannot be answered on the bases of power, but on whether in assessing growth the notion of conditional versus unconditional inference is being raised, Bock (1975).

While the work of Kleinbaum has begun to address the data problems we have in analyzing data in the behavioral sciences, his procedure may lead to spurious test statistics since it depends on the method used to estimate $\hat{\Sigma}_o$ in the construction of the BAN estimator.

References

Bock, R.D. (1963a), Programming univariate and multivariate analysis of variance, *Technometrics*, 5, 95-117.

Bock, R.D. (1963b), Multivariate analysis of variance of repeated measures, *Problems of Measuring Change* (Edited by C.W. Harris), University of Wisconsin Press, Madison, Wisconsin, 85-103.

Bock, R.D. (1975), *Multivariate Statistical Methods in Behavioral Research*, McGraw Hill, New York.

Finn, J. (1972), MULTIVARIANCE: univariate and multivariate analysis of variance, covariance and regression, Version 5, Release 3, June 1976, National Educational Resources Inc., Chicago, Illinois.

Grizzle, J. and D.M. Allen (1969), Analysis of growth and dose response curves, *Biometrics*, 25, 357-381.

Khatri, C.G. (1966), A note on a MANOVA model applied to problems in growth curve, *Annals of the Institute of Statistical Mathematics*, 18, 75-86.

Kleinbaum, D.G. (1970), Estimation and hypothesis testing for generalized multivariate linear models, Institute of Statistics Mimeo Series No. 696, University of North Carolina, Chapel Hill, North Carolina.

Kleinbaum, D.G. (1973), A generalization of the growth curve model which allows missing data, *Journal of Multivariate Analysis*, 3, 117-124.

Lee, Y.H.K. (1974), A note on Rao's reduction of Potthoff & Roy's generalized linear model, *Biometrika*, 61, 349-351.

Olson, Chester L. (1974), Comparative robustness of six tests in multivariate analysis of variance, *Journal of the American Statistical Association*, 69, 894-908.

Potthoff, R.F. and S.N. Roy (1964), A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, 51, 313-326.

Rao, C.R. (1965), The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika*, 52, 447-458.

Rao, C.R. (1966), Covariance adjustment and related problems in multivariate analysis, *Multivariate Analysis* (Edited by P.R. Krishnaiah), Academic Press New York, 87-103.

Rao, C.R. (1967), Least square theory using an estimated disperson matrix and its application to measurement of signals. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics*, University of California Press, Berkeley, California, 355-372.

Rao, C.R. (1972), Recent trends of research work in multivariate analysis, *Biometrics*, 28, 3-22.

Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, 2nd edition, John Wiley and Sons, New York.

Roy, S.N. (1957), *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York.

Roy, S.N. (1964), A report on some results in simultaneous or joint linear (point) estimation, Institute of Statistics, Mimeo Series No. 394, University of North Carolina, Chapel Hill, North Carolina.

Schatzoff, M. (1966), Sensitivity comparisons among tests of the general linear hypothesis, *Journal of the American Statistical Association*, 61, 415-435.

Srivastava, J.N. and Roy, S.N. (1965), Hierarchal and p-block multiresponse designs and their analysis, *Mahalanobis Dedicatory Volume* (Edited by C.R. Rao), Pergamon Press, New York.

Srivastava, J.N. (1966), Some generalizations of multivariate analysis of variance, *Multivariate Analysis* (Edited by P.R. Krishnaiah), Academic Press, New York, 129-145.

Srivastava, J.N. (1967) On the extension of Gauss-Markov theorem to complex multivariate linear models, *Annals of the Institute of Statistical Mathematics*, 19, 417-347.

Srivastava, J.N. (1968), On a general class of designs for multiresponse experiments, *Annals of Mathematical Statistics*, 39, 1825-1843.

Timm, N.H. (1975), *Multivariate Analysis with Applications in Education and Psychology*, Brooks/Cole Publishing Company, Belmont, California.

Trawinski, I.M. and Bargmann, R.E. (1964), Maximum likelihood estimation with incomplete multivariate data. *Annals of Mathematical Statistics*, 35, 647-658.

Tubbs, J. D., T.O. Lewis and B.S. Duran (1975), A note on the analysis of the MANOVA model and its application to growth curves, *Communications in Statistics*, 4, 643-653.